

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2003-249996  
(P2003-249996A)

(43)公開日 平成15年9月5日(2003.9.5)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	マーク* (参考)
H 04 M 1/60		H 04 M 1/60	C 5 D 0 2 0
G 10 L 19/00		H 04 B 3/23	5 D 0 4 5
	21/02	H 04 R 3/02	5 K 0 2 7
H 04 B 3/23		G 10 L 9/18	M 5 K 0 4 6
H 04 R 3/02		9/00	F

審査請求 未請求 請求項の数 7 OL (全 10 頁) 最終頁に続く

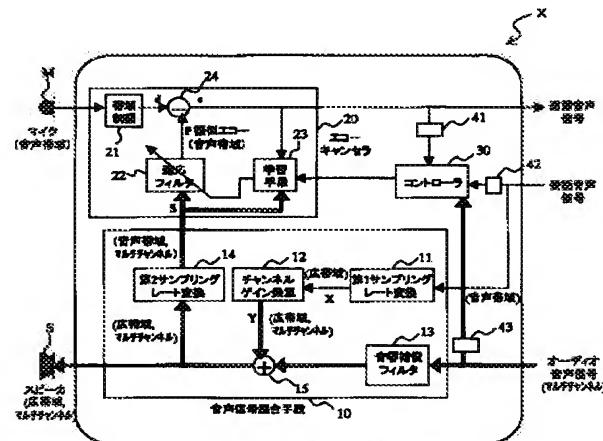
(21)出願番号	特願2002-48183(P2002-48183)	(71)出願人 000001198 株式会社神戸製鋼所 兵庫県神戸市中央区臨浜町二丁目10番26号
(22)出願日	平成14年2月25日(2002.2.25)	(72)発明者 高橋 哲也 兵庫県神戸市西区高塚台1丁目5番5号 株式会社神戸製鋼所神戸総合技術研究所内
		(72)発明者 橋本 裕志 兵庫県神戸市西区高塚台1丁目5番5号 株式会社神戸製鋼所神戸総合技術研究所内
		(74)代理人 100084135 弁理士 本庄 武男
		最終頁に続く

(54)【発明の名称】 音声信号入出力装置

(57)【要約】

【課題】 オーディオ音声信号の出力機能とハンズフリー通話機能とを実現し、オーディオ音声信号と通話信号とが混在する状況化で適切なエコーキャンセルを行う。

【解決手段】 受話音声信号を第1のサンプリングレート変換部11によりオーディオ音声信号のそれと一致させ、さらにチャンネルゲイン乗算部12によりオーディオ音声信号の各チャンネル(L, R等)に対応したチャンネルごとに音量調節した後に、オーディオ音声信号と混合した混合音声信号をスピーカ出力する。さらに、エコーキャンセラ20により、低サンプリングレートに変換した混合音声信号に基づいて、マイクM入力された送話音声信号からエコー除去を行う。エコー除去のパラメータを学習するタイミングは、コントローラ30により、送話及び通話音声信号、オーディオ音声信号のレベルに基づいて制御する。



## 【特許請求の範囲】

【請求項1】 ハンズフリー通話における相手先へ送信する送話音声信号を所定の音声入力手段から入力して所定の通話装置へ出力するとともに、該通話装置から入力した相手先からの受話音声信号を所定の音声出力手段に出力する音声信号入出力装置において、所定のオーディオ音声信号を入力するオーディオ音声信号入力手段と、前記受話音声信号に前記オーディオ音声信号を混合した混合音声信号を前記音声出力手段に出力する音声信号混合手段と、前記混合音声信号に基づいて、前記通話装置へ出力する前記送話音声信号から前記混合音声信号に対応する混入音声信号を除去する混入音声信号除去手段と、を具備してなることを特徴とする音声信号入出力装置。

【請求項2】 前記オーディオ音声信号が1以上のチャンネルから構成されるものであり、前記音声信号混合手段により混合される前記受話音声信号に対して、前記オーディオ音声信号のチャンネルそれぞれに対応した所定の信号処理を施すチャンネル信号処理手段を具備してなる請求項1に記載の音声信号入出力装置。

【請求項3】 前記音声信号混合手段により混合される前記受話音声信号と前記オーディオ音声信号とについて、そのサンプリングレート及び周波数帯域の一方又は両方を一致させるよう信号変換する第1の信号変換手段を具備してなる請求項1又は2のいずれかに記載の音声信号入出力装置。

【請求項4】 前記混入音声信号が除去された前記送話音声信号と前記混合音声信号とにに基づいて、前記混入音声信号除去手段における前記混入音声信号の特定に関するパラメータを学習する学習手段を具備してなる請求項1～3のいずれかに記載の音声信号入出力装置。

【請求項5】 前記学習手段及び前記混入音声除去手段の一方又は両方に用いられる前記混合音声信号及び前記送話音声信号について、そのサンプリングレート及び周波数帯域の一方又は両方を一致させるよう信号変換する第2の信号変換手段を具備してなる請求項4に記載の音声信号入出力装置。

【請求項6】 前記第2の信号変換手段における信号変換後のサンプリングレートが、前記音声信号入力手段から入力される前記送話音声信号、前記通話装置から入力される前記受話音声信号、及び前記オーディオ音声信号の各サンプリングレートの値の最小値に基づいて設定されてなる請求項5に記載の音声信号入出力装置。

【請求項7】 前記送話音声信号、前記受話音声信号、及び前記オーディオ音声信号のいずれか1つ又は複数の信号レベルを検出する信号レベル検出手段と、該信号レベル検出手段による信号レベルに基づいて、前記送話音声信号、前記前記受話音声信号、及び前記オーディオ音声信号のいずれか1つ又は複数の信号レベルを調節する信号レベル調節手段と、を具備してなる請求項1～6の

いずれかに記載の音声信号入出力装置。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、オーディオ音声信号の出力機能とハンズフリー通話機能とを実現する音声信号入出力装置に関するものである。

## 【0002】

【従来の技術】 近年、自動車運転時においても、安全に携帯電話等による通話を行えるようにするため、いわゆるハンズフリー通話を可能とする車載装置が普及しつつある。また、小型化や低コスト化等のため、カーオーディオやカーナビゲーション等の車載装置の一体化、部品の共有化が図られている。このような中で、ハンズフリー通話を実現する装置もカーオーディオ等と一体化することにより、スピーカやマイク、情報処理部等を共有化することが望ましい。また、インターネットにより安価な電話サービスを提供するVoIP (Voiceover IP) の普及にともない、VoIPを用いたハンズフリー通話を実現する装置をゲーム機やテレビ受像機等の情報家電に組み込むことも行われつつある。この場合にも、スピーカやマイク、情報処理部等を共有化することが小型化、低コスト化のために望ましい。一般に、ハンズフリー通話機能を有する通話装置では、スピーカ（音声出力手段）から出力された音声がマイク（音声入力手段）に回り込み、利用者が直接発した音声に混入することによりエコーやハウリングが発生することが知られている。このようなエコーやハウリングを防止するため、混入した音声信号（以下、混入音声信号という）をマイクで入力した音声信号から除去するエコーキャンセル機能を設けることが必要となる。このエコーキャンセル機能を有するものとして、例えば、特開平10-190533号公報、特開平11-284550号公報、特開平11-298382号公報（以下、公報1～3という）等には、前記混入音声信号を学習同定し、相手先からの受話信号に基づいてエコーを除去する適応フィルタを設けた通話装置が示されている。

## 【0003】

【発明が解決しようとする課題】 しかしながら、オーディオや情報家電等にハンズフリー通話機能を設ける場合、スピーカから出力される音声には、相手先からの受話音声信号だけでなく、音楽CD再生やテレビ放送等に係る音声信号（以下、オーディオ音声信号という）も加わることになるが、前記公報1～3に示されるような受話音声信号に基づくエコーキャンセルでは、適切なエコーキャンセルが行えないという問題点があった。また、特開2001-195085号公報（以下、公報4という）には、カーオーディオによる音声が流れる車内で、音声認識装置に入力する音声信号から前記オーディオ音声信号に基づきエコーキャンセルを行う音声認識用オーディオキャンセル装置が示されているが、ハンズフリー

通話機能を設けた場合に、相手先からの受話音声信号に起因するエコーをどのようにキャンセルするかについては何ら示されていない。また、一般に、通話信号（前記送話音声信号及び受話音声信号、即ち、人の声についての音声信号）と前記オーディオ音声信号とは周波数帯域やサンプリングレート等の信号形態が異なるが、この違いをどのように処理してエコーキャンセルを行うかについても、前記公報1～3及び公報4には何ら示されていない。従って、本発明は上記事情に鑑みてなされたものであり、その目的とするところは、オーディオ音声信号の出力機能とハンズフリー通話機能とを実現し、オーディオ音声信号と通話信号とが混在する状況化で適切なエコーキャンセルを行う音声信号入出力装置を提供することにある。

## 【0004】

【課題を解決するための手段】上記目的を達成するため本発明は、ハンズフリー通話における相手先へ送信する送話音声信号を所定の音声入力手段から入力して所定の通話装置へ出力するとともに、該通話装置から入力した相手先からの受話音声信号を所定の音声出力手段に出力する音声信号入出力装置において、所定のオーディオ音声信号を入力するオーディオ音声信号入力手段と、前記受話音声信号に前記オーディオ音声信号を混合した混合音声信号を前記音声出力手段に出力する音声信号混合手段と、前記混合音声信号に基づいて、前記通話装置へ出力する前記送話音声信号から前記混合音声信号に対応する混入音声信号を除去する混入音声信号除去手段と、を具備してなることを特徴とする音声信号入出力装置である。これにより、オーディオ機器や家電製品等とスピーカ（前記音声出力手段）やマイク（前記音声入力手段）等の部品を共有化しながら、オーディオ音声信号の出力機能とハンズフリー通話機能とを実現するに際し、受話音声信号とオーディオ音声信号とを混合した混合音声信号に基づいてエコー除去を行うため、オーディオ音声信号と通話信号とが混在する状況化でもオーディオ音声信号の影響を考慮した適切なエコーキャンセルを行うことが可能となる。

【0005】また、前記オーディオ音声信号が1以上のチャンネルから構成されるものであり、前記音声信号混合手段により混合される前記受話音声信号に対して、前記オーディオ音声信号のチャンネルそれぞれに対応した所定の信号処理を施すチャンネル信号処理手段を具備するものであれば、利用者と前記チャンネルそれぞれに対応した複数のスピーカとの位置関係に応じて音声信号の適切なバランス調整（音量等）を行うことが可能となる。

【0006】また、前記音声信号混合手段により混合される前記受話音声信号と前記オーディオ音声信号について、そのサンプリングレート及び周波数帯域の一方又は両方を一致させるよう信号変換する第1の信号変換手

段を具備すれば、通話信号（前記受話音声信号及び前記送話音声信号）とオーディオ音声信号とのサンプリングレートや周波数帯域に違いがある場合にも適用可能である。

【0007】また、前記混入音声信号が除去された前記送話音声信号と前記混合音声信号とに基づいて、前記混入音声信号除去手段における前記混入音声信号の特定に関するパラメータを学習する学習手段を具備すれば、状況変化に応じた混入音声信号の除去が可能となる。

【0008】また、前記学習手段及び前記混入音声除去手段の一方又は両方に用いられる前記混合音声信号及び前記送話音声信号について、そのサンプリングレート及び周波数帯域の一方又は両方を一致させるよう信号変換する第2の信号変換手段を具備するものも考えられ、さらに、前記第2の信号変換手段における信号変換後のサンプリングレートが、前記音声信号入力手段から入力される前記送話音声信号、前記通話装置から入力される前記受話音声信号、及び前記オーディオ音声信号の各サンプリングレートの値の最小値に基づいて設定すれば、混入音声除去やその学習の演算負荷を大幅に低減でき、比較的性能の低い、安価なDSPやCPU等で混入音声の除去やその学習を実現できる。

【0009】また、前記送話音声信号、前記受話音声信号、及び前記オーディオ音声信号のいずれか1つ又は複数の信号レベルを検出する信号レベル検出手段と、該信号レベル検出手段による信号レベルに基づいて、前記送話音声信号、前記前記受話音声信号、及び前記オーディオ音声信号のいずれか1つ又は複数の信号レベルを調節する信号レベル調節手段と、を具備すれば、例えば、通話中は前記オーディオ音声信号のレベルを下げる等の調節により、混入音声除去の学習が不十分な場合等においても、送話音声信号へのエコーの混入やハウリングの発生を防止することができる。

## 【0010】

【発明の実施の形態】以下添付図面を参照しながら、本発明の実施の形態及び実施例について説明し、本発明の理解に供する。尚、以下の実施の形態及び実施例は、本発明を具体化した一例であって、本発明の技術的範囲を限定する性格のものではない。ここに、図1は本発明の実施の形態に係る音声信号入出力装置Xの構成を表すブロック図、図2は本発明の実施例に係る音声信号入出力装置Y1の構成を表すブロック図、図3は本発明の実施例に係る音声信号入出力装置Y2の構成を表すブロック図である。

【0011】まず、図1を用いて本発明の実施の形態に係る音声信号入出力装置Xについて説明する。本音声信号入出力装置Xは、例えば、マイクMとスピーカSが設けられたカーオーディオやテレビ受像機等のオーディオ機器に組み込まれ、携帯電話機や通常の電話機等の通話装置と接続されることにより、いわゆるハンズフリー通

話機能を実現するものである。即ち、オーディオ機器が備えるマイクM（前記音声信号入力手段の一例）により送話音声信号を入力し、これを電話機等の通話装置に出力することによって相手先へ送信するとともに、相手先から受信した受話音声信号を電話機等から入力し、これをオーディオ機器のスピーカS（前記音声信号出力手段の一例）に出力して音声出力するものである。このとき、スピーカSには、前記受話音声信号にオーディオ機器より出力されるオーディオ音声信号（ラジオ放送の受信音声やCD音楽の再生音等の音声信号）を混合した

（加えた）混合音声信号が出力される。これにより、オーディオ機器のスピーカSが、通話用のスピーカとしても兼用（共有化）されることになる。図1に示すように、本音声信号入出力装置Xは、前記受話音声信号と前記オーディオ音声信号とを混合し、混合された前記混合音声信号を不図示のD/A変換器及びアンプを介してスピーカSに出力する音声信号混合手段10と、前記混合音声信号に基づいて、マイクMから入力して前記通話装置へ出力する前記送話音声信号から前記混合音声信号に対応する混入音声信号を除去するエコーキャンセラ20と、前記オーディオ音声信号と前記通話装置へ出力される前記送話音声信号とを入力し、これらに基づいて前記エコーキャンセラ20が具備する所定の学習手段23

（後述）が学習を行うべきタイミングを制御するコントローラ30とを具備している。このように、前記エコーキャンセラ20が、前記受話音声信号と前記オーディオ音声信号とが混合された前記混合音声信号に基づいてエコー除去を行うため、前記オーディオ音声信号と前記受話音声信号とが混在する状況化で適切なエコーキャンセルを行うことが可能となる。

【0012】次に、前記音声信号混合手段10について説明する。一般に、音楽CD等の前記オーディオ音声信号は、サンプリングレートが4.4.1kHzであり、理論的には22.05kHzという高い周波数帯域まで含むことが可能な広帯域な音声信号である。一方、電話機等の通話装置で用いられる通話信号（前記受話音声信号及び送話音声信号）は、一般に、サンプリングレートが8kHzであり、4kHz程度の周波数帯域しか含まざることができない。このようにサンプリングレートや周波数帯域の異なる2つの音声信号を混合するために、前記音声信号混合手段10は、前記受話音声信号及び前記オーディオ音声信号について、混合前にそれらのサンプリングレートを一致させるよう変換するための第1のサンプリングレート変換部11を具備している。図1に示す前記第1のサンプリングレート変換部11は、前記受話音声信号のサンプリングレートを前記オーディオ音声信号のそれ（4.4.1kHz）に一致させるよう変換するものである。変換後のサンプリングレートが、前記2つの音声信号のうちの高い方の値となるように構成されているのは、スピーカSへ出力される音声信号に音質の

低下が生じないようにするためである。前記第1のサンプリングレート変換部11としては、例えば、変換前のデジタル信号をアップサンプリングした信号を低域通過フィルタに通した後、目標のサンプリングレートになるようにダウンサンプリングするものが知られている。具体的には、8kHzから44.1kHzへの変換を行う場合、まず、8kHzの前記受話音声信号を44.1倍のサンプリングレートとし、44.1サンプルごとに元の8kHzの音声信号と同じ値を残し、その他の値が0（ゼロ）にした信号を作るいわゆるアップサンプリング処理を施す。この信号を44.1kHzにダウンサンプリングする前に、8kHzでは本来存在しない4kHzを越える周波数成分を取り除くための低域通過フィルタの処理を施す。このフィルタ処理により得られた信号を80サンプルおきに間引いた信号を作ることにより44.1kHzの音声信号が得られる。もちろん、サンプリングレートの変換手段としては、データの補間を用いるものなど各種存在するので、前記第1のサンプリングレート変換部11は、他の手段によるものであってもかまわない。

【0013】また、前記オーディオ音声信号は、いわゆるモノラル方式のように1チャンネルで構成される場合は少なく、一般に、いわゆるステレオ方式の2チャンネル、或いはサラウンド方式の5.1チャンネル等のように複数のスピーカSそれぞれに対応した複数チャンネルで構成される場合が多い。一方、前記通話装置で用いられる前記通話信号は、モノラル方式であることが一般的である。例えば、カーオーディオ等のように、主な聴取者である運転者が座る運転席と複数のスピーカSとの位置関係が固定されている場合には、運転席において最も音量バランス（左右バランス等）が良くなるように各スピーカSの音量レベル、即ち、前記オーディオ音声信号の各チャンネルの信号レベルを個別に補正（信号処理の一例）することが有効である。これを実現するため、前記音声信号混合手段10は、前記受話音声信号を前記オーディオ音声信号の各チャンネルに対応するチャンネルごとに信号レベルのゲイン補正を行うチャンネルゲイン乗算部12（前記チャンネル信号処理手段の一例）を具備している。これにより、例えば、ステレオ方式の場合には左右の2チャンネル、サラウンド方式の場合には5.1チャンネル（5+1チャンネル）それぞれに対応する信号のゲインが補正（それぞれ個別に乗算補正）される。さらに、ゲインの乗算補正だけでなく、車内の音響特性を考慮して各チャンネルごとに所定の特性のフィルタ処理（信号処理の一例）を施すようにしてよい。

【0014】以上のように、前記第1のサンプリングレート変換部11によりサンプリングレートが前記オーディオ音声信号のそれと一致するよう変換され、チャンネルごとに所定の処理が施された前記受話音声信号に対し、所定の音響補償フィルタ13により所定の音響処理

がなされた前記オーディオ音声信号を加算器15により混合し(加え)，混合後の混合音声信号を不図示のA/D変換器及びアンプを介してスピーカSに出力する。もちろん、前記チャンネルゲイン乗算部12と同様に、前記オーディオ音声信号や前記混合音声信号についてもチャンネルごとに信号処理(ゲイン乗算やフィルタ処理等)を行う手段を設けることも考えられる。また、前記混合音声信号は、前記音声信号混合手段10の具備する第2サンプリングレート変換部14により、そのサンプリングレートが変換されて前記エコーキャンセラ20に出力される。前記第2のサンプリングレート変換部14による変換後のサンプリングレートの値は、前記通話装置へ出力される前記送話音声信号のサンプリングレート(8kHz)の値となるように設定されている。

【0015】次に、前記エコーキャンセラ20について説明する。前記エコーキャンセラ20は、帯域制限部21、適応フィルタ22、学習手段23、及び減算器24を具備している。前記帯域制限部21は、マイクMにより入力された前記送話音声信号の周波数帯域を所定のフィルタにより、前記通話装置へ出力される前記送話音声信号の周波数帯域(4kHz)に制限(変換)するものである。前記帯域制限部21及び前記第2のサンプリングレート変換部14により、前記第2の信号変換手段の一例が構成されている。前記適応フィルタ22は、前記音声信号混合手段10から入力した低サンプリングレート

\*ト(8kHz)に変換後の前記混合音声信号に基づいて、スピーカSからマイクMに回り込む音声に相当する擬似エコー(前記混合音声信号に対応する前記混入音声信号の一例)を出力するものである。前記減算器24は、前記擬似エコーを、前記帯域制限部21により帯域制限された前記送話音声信号(前記通話装置へ出力される前記送話音声信号)から減算器24によって除去することにより、スピーカSからマイクMに回り込むエコーを除去するものである。前記適応フィルタ22及び前記減算器24により、前記混入音声信号除去手段の一例が構成されている。前記学習手段23は、前記音声信号混合手段10から入力した低サンプリングレート(8kHz)の前記オーディオ音声信号と、前記擬似エコーが除去された前記送話音声信号に基づいて、前記適応フィルタ22のフィルタ係数(前記擬似エコーを求める(特定する)ためのパラメータの一例)を自動学習するものである。該学習手段23の学習方法としては、LMSアルゴリズムや学習同定アルゴリズム等の手法が知られている。例えば、学習同定アルゴリズムでは、最大Nサンプル時間までの長さのエコーを除去する場合、前記擬似エコーを $p(i)$ とすると、 $p(i)$ は過去Nサンプル分のL(左)側スピーカの音声出力 $S_L(i)$ 、及びR(右)側スピーカの音声出力 $S_R(i)$ の線形和として次の(1)式で推定する。

【数1】

$$\begin{bmatrix} S_L(i) \\ S_L(i-1) \\ \vdots \\ S_L(i-N+1) \\ S_R(i) \\ S_R(i-1) \\ \vdots \\ S_R(i-N+1) \end{bmatrix} \cdots (1)$$

$$p(i) = \{ K_L, K_L, \dots, K_L, K_R, K_R, K_R, \dots \}$$

ここで、 $K_L$ 、 $K_R$ は適応フィルタのフィルタ係数である。この(1)式の時刻 $i$ における前記フィルタ係数 $K_L$ 、 $K_R$ からなる1行ベクトルを $H(i)$ 、同スピーカの音声出力 $S_L(i)$ 、 $S_R(i)$ からなる1列ベクトルを $S(i)$ とし、前記擬似エコー $p(i)$ の誤差を $e(i)$ (時刻 $i$ におけるマイクMからの実際の※

$$H(i+1) * H(i) + \alpha \frac{S(i)}{\|S(i)\|^2} * (i) \cdots (2)$$

(2)式において、 $\alpha$ は学習の収束スピードを調節するパラメータであり、各種の設定方法が知られている。前記適応フィルタ22及び前記学習手段23については、前記公報1～3等にも示されている。本音声信号入出力装置Xでは、図1に示すように、マイクMから入力した前記送話音声信号から前記擬似エコーを除去した後の信号を前記誤差 $e(i)$ として用いている。

【0016】前述したように、本音声信号入出力装置X

※音声入力と前記擬似エコー $p(i)$ との差分)とする。学習同定アルゴリズムでは、次の(2)式に示すように、誤差 $e(i)$ に応じて前記フィルタ係数を修正(学習)を行う。

【数2】

では、前記適応フィルタ22及び前記学習手段23に入力される音声信号のサンプリングレートは、前記帯域制限部21及び前記第2のサンプリングレート変換部14によって小さい値、即ち、マイクMから入力される前記送話音声信号、スピーカSへ出力される前記混合音声信号、及び前記通話装置へ出力される前記送話音声信号の各サンプリングレートのうち最小の値となるように変換される。これにより、DSPやCPU等で実現される前

記適応フィルタ22や前記学習手段23にかかる処理の負荷を大幅に低減することができ、比較的性能の低い、安価なDSPやCPU等で前記適応フィルタ22や前記学習手段23を実現することが可能となる。一般に、取り扱う信号のサンプリングレートが大きくなるにつれて必要な演算量が加速度的に増大するので、サンプリングレートを小さくした信号に基づき前記適応フィルタ22や前記学習手段23の処理を行うことによる負荷低減効果は大きい。一方、このようにサンプリングレートを低くすると、前記送話音声信号の音声品質を低下させることになるが、そもそも通話用としては前記オーディオ音声信号などの音声品質を必要としないため、このようにしても何ら問題はない。

【0017】次に、前記コントローラ30について説明する。前述したように、前記学習手段23は、前記擬似エコーを除去した後の前記送話音声信号を前記誤差e

(1)として用いているため、話者が発した音声と、スピーカSから回りこんだエコーとが混在してマイクMに入力されるダブルトークの状況化で学習を行うと、話者の発した音声が外乱となるため適切な学習が行えない。そこで、話者が声を発していないタイミングで学習を行う必要がある。前記コントローラ30は、この学習のタイミングを、前記擬似エコーを除去した後の前記送話音声信号の信号レベルに基づいて制御する。該信号レベルは所定の信号レベル検出手段41により検出される。例えば、前記送話音声信号の信号レベルが所定レベル以下である場合に前記ダブルトークの状態でないと判断し、該受話状態である場合にのみ前記学習手段23による前記適応フィルタ22のパラメータの学習を行うよう制御する等が考えられる。これにより、話者の音声が混入しない適切なタイミングでエコーの学習が可能となる。もちろん、前記受話音声信号や、前記オーディオ音声信号の信号レベルを検出し、これらに基づいて学習タイミングを制御するようにしてもよい。例えば、前記受話音声信号のレベルが、前記送話音声信号のレベルより大きい場合に前記受話状態であると判断することや、前記受話状態であって、かつ前記オーディオ音声信号のレベルが大きい場合にのみ学習する等が考えられる。

【0018】

【実施例】次に、図2を用いて、前記音声信号入出力装置Xの応用例である音声信号入出力装置Y1について説明する。前述したように、前記音声信号入出力装置Xでは、前記エコーキャンセラ20によりスピーカSからマイクMに回り込むエコーを除去するが、前記学習手段23による学習が不十分な場合等にはエコー除去が不十分となる場合もある。このため、前記エコーキャンセラ20によるエコー除去に加えて、前記通話装置を介しての通話状態に応じて、各音声信号の信号レベル(音量レベル)を調節することが望ましい。これに対応するため、本音声信号入出力装置Y1は、前記音声信号入出力装置

Xにおける前記コントローラ30に新たな制御機能を付加したものであり、その他の構成及び機能については、前記音声信号入出力信号装置Xと同じである。即ち、本音声信号入出力装置Y1は、前記音声信号入出力装置Xの構成に加え、前記通話装置へ出力される前記送話音声信号、前記音声信号混合手段10に入力される前記受話音声信号及び前記オーディオ音声信号の音量レベル(信号レベル)それぞれを検出する音量レベル検出手段41、42、43及び、同音量レベルそれぞれを調節する音量ゲイン調節手段51、52、53を具備している。ここで、前記送話音声信号についての前記音量レベル検出手段41は、前記エコーキャンセラ20により前記擬似エコーが除去された後の信号レベルを検出することが望ましい。前記コントローラ30は、前記音量レベル検出手段41、42、43それぞれによって検出された音量レベルに基づいて前記音量ゲイン調節手段51、52、53それぞれを例えば以下のように制御する。まず、前記送話音声信号のレベルが前記受話音声信号のレベルよりも大きい場合は、送話状態であると判断し、前記受話音声信号に係る前記音量ゲイン調節手段52のゲインを下げる。逆に、前記受話音声信号のレベルが前記送話音声信号のレベルよりも大きい場合は受話状態であると判断し、前記送話音声信号に係る前記音量ゲイン調節手段51のゲインを下げる。さらには、前記送話音声信号又は前記受話音声信号のいずれかのレベルが所定レベル以上である場合には、通話状態(前記送話状態又は受話状態)であると判断し、前記オーディオ音声信号に係る前記音量ゲイン調節手段53のゲインを下げる。以上のような制御を行うことにより、前記学習手段23による学習が不十分な場合でも、前記送話音声信号にエコーが混入することを防止できるとともに、前記送話音声信号と前記受話音声信号とが同時に発生することによるハウリングの発生防止にもつながる。

【0019】次に、図3を用いて、前記音声入出力装置Xの別の応用例である音声入出力装置Y2について説明する。本発明は、音声認識機能を有するカーナビゲーション装置やカーオーディオ、携帯電話やVoIPによる通話接続等への適用が考えられるが、これらの装置それぞれの有する機能の一部を前記音声入出力装置Xに付加して一体に統合したものが、本音声入出力装置Y2である。図3に示すように本音声入出力装置Y2は、前記音声信号入出力装置Xの構成に加え、ノイズサプレッサ61、音声認識手段62、送話音声符号化手段63、受話音声復号化手段64、通信プロトコル処理手段65、音声合成手段66、オーディオ信号復号化手段67、及び加算器68を具備している。前記ノイズサプレッサ61は、前記エコーキャンセラ20から出力される前記送話音声信号から車の走行音等、所定の周波数帯域のノイズを除去するものである。前記音声認識手段62は、一般に、カーナビゲーション装置等に搭載されるものであ

り、前記ノイズサプレッサ61の出力する音声信号に基づき、音声認識を行い、カーナビゲーション装置に対する制御コマンド等の音声認識結果を出力するものである。これにより、前記エコーチャンセラ20を前記送話音声信号からのエコー除去だけでなく、前記音声認識手段62に対する音声信号からのエコー除去にも兼用（共用）される。前記送話音声符号化手段63及び前記受話音声復号化手段64は、携帯電話やVoIPによる通話装置との間で入出力する前記送話音声信号及び前記受話音声信号について、所定の符号化（エンコード）及び復号化（デコード）をおこなうものである。一般に、携帯電話やVoIP等で送受信される通話信号は、VSELP, PS1-CELP, AMR, G.729, G.723.1等の規格で圧縮（符号化）されている。前記送話音声符号化手段63及び前記受話音声復号化手段64は、これら各規格に応じた符号化及び復号化を行う。この送話音声符号化手段63により符号化する音声信号に車の走行音等のノイズが混入していると、符号化（圧縮）によって著しく音声品質の劣化が生じるので、前記ノイズサプレッサ61によるノイズ除去が、音声品質を確保する上で有効である。前記通信プロトコル処理手段65は、例えば、VoIPにおける RTPやG.723.1, SIP等、音声信号の符号化や復号化と同時にを行うことが効率的である通信のプロトコル処理を行うものである。前記音声合成手段66は、カーナビゲーション装置等で行われる利用者への音声ガイダンス等の合成音声を出力するものである。該合成音声は、前記加算器68によって前記音声信号混合手段10に入力される前記受話音声信号に加算（混合）される。これにより、音声ガイダンス等の合成音声も前記エコーチャンセラ20が用いる前記混合音声信号に含まれることになるので、前記送話音声信号に音声ガイダンス等のエコーが含まれることを防止できる。前記オーディオ信号復号化手段67は、MP3, WMA, TwinVQ, MPEG2-AA等の圧縮規格によって符号化されたデジタルオーディオ信号（前記オーディオ音声信号の一例）を復号化するものである。

【0020】図3に示した各構成要素を1チップのDSP等に一体化して統合することは、単に複数の機器を組み合わせた1チップ化により小型・軽量化が図れるだけでなく、通話品質悪化を防止する効果をも奏する。即ち、各機器を別の装置に搭載して接続し、それぞれの間で音声信号を入出力する構成とした場合、機器間で通信を行うことにより生じる音声信号の入出力のタイミングの違いを緩衝させたり、信号の処理待ち等のためにバッファが必要となるが、このようなバッファに入出力信号が一時滞留すると信号処理の遅延が生じる。例えば、送話側では、前記エコーチャンセラ20、前記ノイズサプレッサ61、前記送話音声符号化手段63、及び前記通信プロトコル処理手段65それぞれの間の信号入出力に

おいて遅延が生じ、受話側では、前記通信プロトコル処理手段65、前記受話音声復号化手段64、及び前記音声信号混合手段10それぞれの間の信号入出力において遅延が生じる。この遅延により、例えば一方の話者の問い合わせに対して他方の話者の応答が遅れて聞こえてくる等、円滑な通話が阻害される。一方、図3に示す構成要素を1チップ化して統合することにより、同一モリ上のデータ（音声信号）に対して各種処理を行えるようになり、音声信号の入出力に遅延が発生せず、通話品質の悪化を防止できる。さらに、処理の遅延がなくなることにより、利用者の音声を認識してガイダンス（合成音声）を出力したりオーディオ機器を動作させたりする等の応答の速度が向上するという効果も奏する。また、前記音声信号入出力装置Y1と同様に音量調節機能を付加することにより、例えば前記音声認識手段62の動作中はスピーカSへの出力信号を減衰させる等、全体として最適な制御が可能となる。

## 【0021】

【発明の効果】以上説明したように、本発明によれば、オーディオ機器や家電製品等とスピーカやマイク等の部品を共有化しながら、オーディオ音声信号の出力機能とハンズフリー通話機能とを実現するに際し、受話音声信号とオーディオ音声信号とを混合した混合音声信号に基づいてエコー除去を行うため、オーディオ音声信号と通話信号とが混在する状況でもオーディオ音声信号の影響を考慮した適切なエコーチャンセルを行うことが可能となる。また、オーディオ音声信号の有する1以上のチャネルに対応させて受話音声信号の処理を行うことにより、利用者と複数のスピーカとの位置関係に応じて音声信号の適切なバランス調整を行うことが可能となる。また、サンプリングレートや周波数帯域についての信号変換手段を具備することにより、通話信号とオーディオ音声信号とのサンプリングレートや周波数帯域に違いがある場合にも適用可能である。また、混入音声（エコー）の除去やその学習に用いる信号のサンプリングレートを、通話信号及びオーディオ音声信号のサンプリングレートの最小値に基づいて設定することにより、混入音声除去やその学習の演算負荷を大幅に低減でき、比較的性能の低い、安価なDSPやCPU等で混入音声の除去やその学習を実現できる。また、送話音声信号、受話音声信号、オーディオ音声信号について、これらの信号レベルの検出結果に基づいてその信号レベルを調節することにより、エコー除去の学習が不十分な場合等においても、送話音声信号へのエコーの混入やハウリングの発生を防止することができる。また、本発明を、音声信号の符号化や復号化、音声認識、通信プロトコル処理、音声合成等の各手段と統合した形で実施すれば、信号の遅延が防止できるため、信号遅延に起因する通話品質の悪化を防止できるとともに、応答速度の向上が図れる。

## 【図面の簡単な説明】

【図1】本発明の実施の形態に係る音声信号入出力装置Xの構成を表すブロック図。

【図2】本発明の実施例に係る音声信号入出力装置Y1の構成を表すブロック図。

【図3】本発明の実施例に係る音声信号入出力装置Y2の構成を表すブロック図。

【符号の説明】

- 1 0…音声信号混合手段
- 1 1…第1のサンプリングレート変換部
- 1 2…チャンネルゲイン乗算部
- 1 3…音響補償フィルタ
- 1 4…第2のサンプリングレート変換部
- 1 5…加算器
- 2 0…エコー・キャンセラ
- 2 1…帯域制限部
- 2 2…適応フィルタ

\* 2 3…学習手段

2 4…減算器

3 0…コントローラ

4 1, 4 2, 4 3…音量レベル検出手段

5 1, 5 2, 5 3…音量ゲイン調節手段

6 1…ノイズサプレッサ

6 2…音声認識手段

6 3…送話音声符号化手段

6 4…受話音声復号化手段

10 6 5…通信プロトコル処理手段

6 6…音声合成手段

6 7…オーディオ信号復号化手段

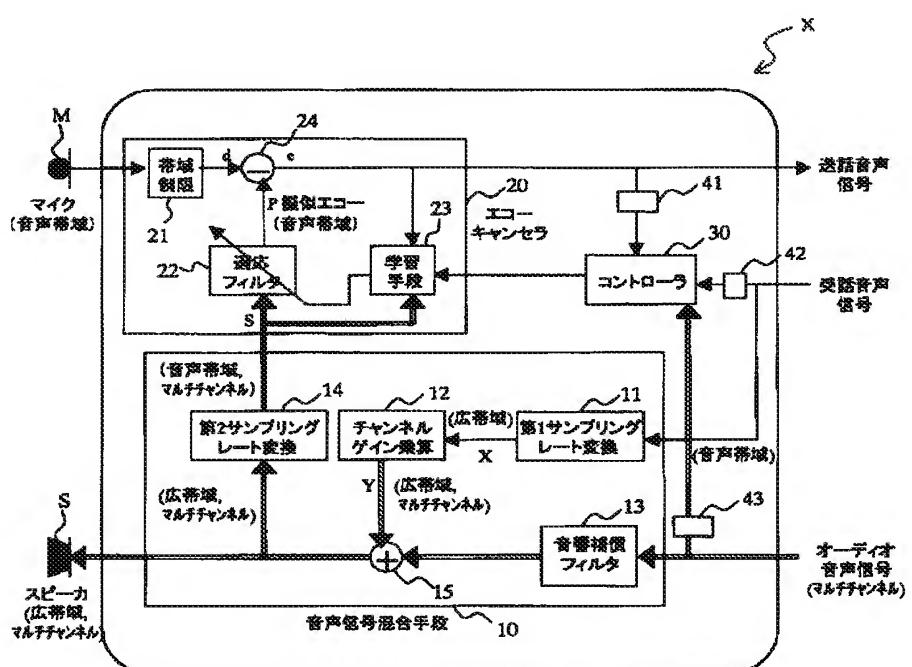
6 8…加算器

M…マイク

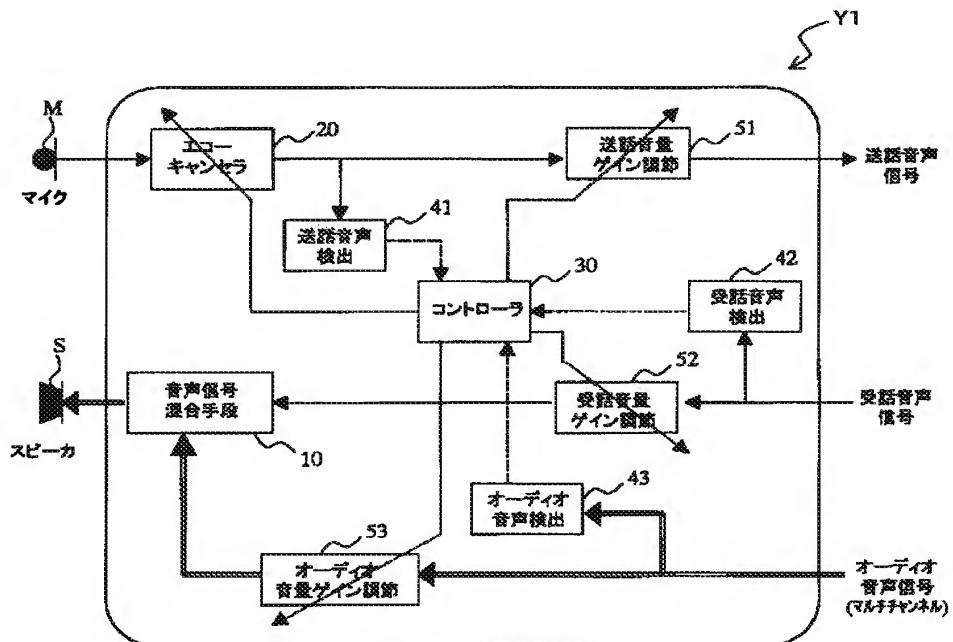
S…スピーカ

\*

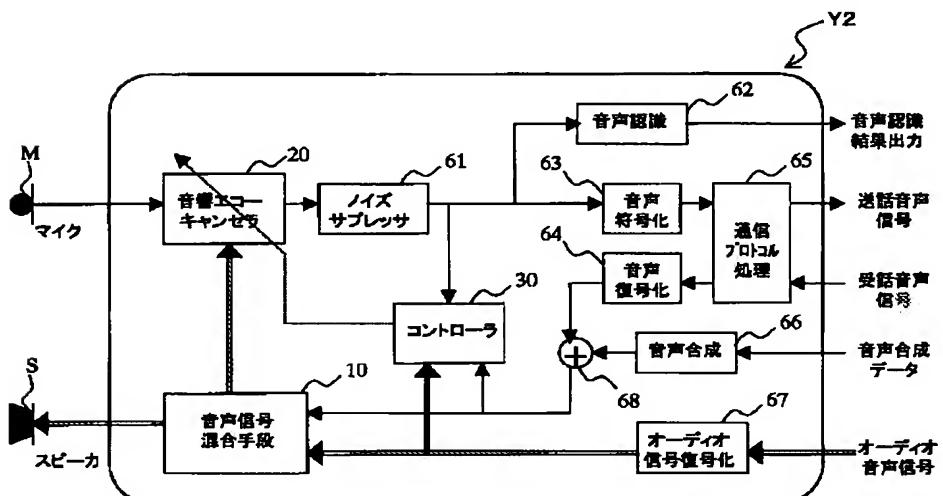
【図1】



【図2】



【図3】



フロントページの続き

(72)発明者 下田 敏章 F ターム(参考) 5D020 CC05 CC06  
兵庫県神戸市西区高塚台1丁目5番5号 5D045 AB26 DA20  
株式会社神戸製鋼所神戸総合技術研究所内 5K027 DD07 DD10 HH03  
(72)発明者 稚方 孝之 5K046 BB01 CC28 HH02 HH19 HH78  
兵庫県神戸市西区高塚台1丁目5番5号 HH79  
株式会社神戸製鋼所神戸総合技術研究所内